Remote Sensing of Environment 208 (2018) 42–62

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A robust adaptive spatial and temporal image fusion model for complex land surface changes

Yongquan Zhao[a], Bo Huang[a,b,*,1], Huihui Song[c]

[a] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, China
[b] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China
[c] Jiangsu Key Laboratory of Big Data Analysis Technolog, Nanjing University of Information Science and Technology, Nanjing 210044, China

ARTICLE INFO

Keywords:
Spatiotemporal fusion
Non-shape change
Shape change
NL-LR
Image super-resolution

ABSTRACT

Spatial and temporal satellite image fusion (STIF) has provided a feasible alternative for generating imagery with both high spatial and temporal resolution, thus expanding the applications of existing satellite sensors. However, a critical challenge confronting the further development of STIF is to systematically and robustly address complex land surface changes, which include land cover changes without shape changes (e.g., crop rotation) and land cover changes with shape changes (e.g., urban expansion), in addition to conventional land surface changes (e.g., phenological changes of vegetation). This paper presents the Robust Adaptive Spatial and Temporal Fusion Model (RASTFM) to tackle this challenge with one prior pair of MODIS-Landsat images. In RASTFM, land surface changes are reorganized into non-shape changes (including phenological changes and land cover changes without shape changes) and shape changes (i.e., land cover changes with shape changes), which are handled differently. However, both non-shape changes and shape changes are predicted through a Non-Local Linear Regression (NL-LR) of the subject pixel's similar neighbors. A regression based high-pass modulation is also performed as a post-processing step to improve both the spatial details and spectral fidelity of the predicted Landsat image. Unlike other STIF models (e.g., the Spatial and Temporal Adaptive Reflectance Fusion Model, STARFM), RASTFM can find similar neighboring pixels more precisely through a non-local searching strategy and derives the weights of similar neighbors more rigorously via a linear regression model. As both non-shape and shape changes are treated based on the regression of similar neighboring pixels, the land surface changes are processed in a unified manner. Experiments that use one simulated and three actual MODIS-Landsat datasets featured by different types of land surface changes were conducted to demonstrate the performance of RASTFM. Comparisons with the state-of-the-art STIF models, including weighted function, unmixing and dictionary-learning methods, show that NL-LR based RASTFM can capture the land surface changes in various landscapes more accurately and robustly in a unified manner, thereby facilitating the continuous and detailed monitoring of complex and diverse land surface dynamics.

1. Introduction

High spatial resolution images with frequent coverage are of great significance for many applications at the global or regional scale, such as land cover/land use mapping and change detection (Acerbi-Junior et al., 2006; Roy et al., 2014; Townshend et al., 2012; Xian and Crane, 2005), disturbance events mapping (Hilker et al., 2009), quantitative crop growth monitoring (Singh, 2011), vegetation phenological change monitoring (Bhandari et al., 2012), detailed seasonal variation reconstruction of biophysical parameters (Zhang et al., 2014), land surface temperature monitoring (Weng et al., 2014), and mapping daily evapotranspiration at field or continental scales (Cammalleri et al., 2013; Kustas et al., 2011). To date, although Earth Observation (EO) has made great breakthroughs in obtaining remote sensing data with high spatial and temporal resolution from multi-platform satellites, such as the new launch of the Satellite Pour l'Observation de la Terre (SPOT), Landsat, Geostationary Operational Environmental Satellites (GOES), Meteosat Second Generation (MSG), Multifunctional Transport Satellites (MTSAT), Sentinal-2 satellites, and the development of the China High-resolution Earth Observation System (CHEOS), current satellite sensors still have to compromise between spatial and temporal resolution because of hardware technology limitations and budget

* Corresponding author.
E-mail address: bohuang@cuhk.edu.hk (B. Huang).
[1] Co-first author.

https://doi.org/10.1016/j.rse.2018.02.009
Received 15 May 2017; Received in revised form 23 November 2017; Accepted 8 February 2018

constraints (Gevaert and García-Haro, 2015; Huang and Song, 2012; Song and Huang, 2013; Zhu et al., 2010; Zhu et al., 2016). Moreover, many historical EO data entail the use of STIF for studies on long-term and detailed land surface dynamics that require dense historical time-series satellite images with high spatial resolution. Hence, STIF is a feasible and cost-effective method of obtaining remote sensing images with both high spatial and temporal resolution simultaneously to promote applications of current EO data. (Gao et al., 2006; Huang and Song, 2012; Zhu et al., 2016).

Image fusion can integrate multi-source images to obtain more information than what can be derived from a single data source (Van Genderen and Pohl, 1994). This process can be categorized into three processing levels: pixel-level, feature-level and decision-level (Polh and Van Genderen, 1998). Pixel-level image fusion methods are usually very efficient and easy to implement but highly sensitive to mis-registration (Li et al., 2008); however, feature- and decision-level methods have much stronger tolerance to the imperfection of image registration accuracy (Karali et al., 2015). The primary goal of STIF is to predict unavailable or missing high-spatial-resolution (hereafter referred to as "high-resolution") images, which is caused by the compromise between spatial and temporal resolution, cloud cover, or other interference factors, such as the SLC-off problem of Landsat-7 Enhanced Thematic Mapper Plus (ETM+), by capturing land surface temporal changes from low-spatial-resolution (hereafter referred to as "low-resolution") images but with frequent coverage (e.g., MODerate resolution Imaging Spectroradiometer (MODIS) imagery) as much as possible and by taking full advantage of the spatial details of prior high-resolution images but with less frequent coverage (e.g., Landsat imagery). The combination of existing satellite observations and fused data can achieve land surface dynamics monitoring at higher spatiotemporal resolution that is heretofore inaccessible (Gao et al., 2015). To this end, a good STIF method should reconstruct spatial details, reduce spectral distortion, and exclude the effects of possible disturbances. Although STIF can produce synthetic satellite observations with high spatiotemporal resolution, it still relies on the availability of actual observations (Gao et al., 2015). Hence, STIF is mostly useful for bridging the gaps between spatial and temporal resolution of satellite sensors, but cannot replace actual satellite missions (Gao et al., 2015).

To date, many STIF algorithms have been devised to predict different types of land surface temporal changes, including phenological changes and land cover changes. A general comparison in terms of prediction ability and processing level among several representative STIF methods is shown in Table 1. As summarized by Zhu et al. (2016), most of the weighted function based and unmixing based STIF methods can only predict phenological changes by processing at the pixel level, while all of the dictionary-pair learning methods can capture phenological and land cover changes together by processing at the feature level.

Among the STIF methods featuring only the ability to predict phenological changes, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), proposed by Gao et al. (2006), is the most popular method within the remote sensing community (Emelyanova et al., 2013) and performs well in preserving the spatial details of prior high-resolution images, such as Landsat images (Singh, 2011). Several improved models based on STARFM have since been developed, such as the Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH) (Hilker et al., 2009), the Enhanced STARFM (ESTARFM) (Zhu et al., 2010), the operational STARFM data fusion framework (Wang et al., 2014), and the spatio-temporal integrated temperature fusion model (STITFM) (Wu et al., 2015). Roy et al. (2008) presented a semi-physical fusion approach that considered the directional dependence of surface reflectance described by the Bi-directional Reflectance Distribution Function (BRDF) using the MODIS BRDF/Albedo product. Zurita-Milla et al. (2009) proposed a constrained unmixing method to produce synthetic image series with Landsat-like spatial, and MERIS-like (MEdium Resolution Imaging Spectrometer) spectral and temporal resolution. Wu et al. (2012) developed the Spatial Temporal Data Fusion Approach (STDFA) by performing unmixing at every prior time point to obtain the corresponding surface reflectance changes and then calculating the final predictions based on these changes and a prior high spatial resolution image. Huang et al. (2013) proposed a unified fusion method based on Bayesian data fusion theory that aims to perform spatiotemporal fusion and spatial-spectral fusion in the same process. Amorós-López et al. (2013) added a regularization term to the unmixing cost function to prevent the spectral shape of the derived endmembers from excessively differing from the predefined endmembers spectra. Gevaert and García-Haro (2015) proposed the Spatial and Temporal Reflectance Unmixing model (STRUM) by conducting unmixing on low-resolution images and introducing Bayesian theory to constrain the estimated endmembers. STIF methods that consider using intermediate spatial resolution data as auxiliary data to improve the prediction precision of STIF and reduce the uncertainty in predicting high-resolution images on a desired date have also been proposed, e.g., using 15 m Landsat panchromatic images in the fusion of 30 m Landsat and 10 m Sentinel-2 images (Wang et al., 2017a) and 250 m MODIS images in the fusion of 500 m MODIS and 30 m Landsat images (Wang et al., 2017b). Moreover, Wang and Atkinson (2017) designed a Fit-FC method to fuse the spatial and temporal resolution from Sentinel-2 Multispectral Imager (MSI) and Sentinel-3 Ocean and Land Color Instrument (OLCI) sensors to create nearly daily Sentinel-2 images.

For STIF methods enabling the prediction of both phenological and

Table 1
A general comparison among several representative STIF methods.

STIF method	Phenological change	Land cover change	Processing level
STARFM (Gao et al., 2006)	√	N/A	Pixel-level
Semi-physical model (Roy et al., 2008)	√	N/A	Pixel-level
STAARCH (Hilker et al., 2009)	√	N/A	Pixel-level
Constrained unmixing (Zurita-Milla et al., 2009)	√	N/A	Pixel-level
ESTARFM (Zhu et al., 2010)	√	N/A	Pixel-level
STDFA (Wu et al., 2012)	√	N/A	Pixel-level
Regularized unmixing (Amorós-López et al., 2013)	√	N/A	Pixel-level
Unified fusion (Huang et al., 2013)	√	N/A	Pixel-level
Operational STARFM (Wang et al., 2014)	√	N/A	Pixel-level
STRUM (Gevaert and García-Haro, 2015)	√	N/A	Pixel-level
SPSTFM (Huang and Song, 2012)	√	√	Feature-level
One-pair learning (Song and Huang, 2013)	√	√	Feature-level
U-STFM (Huang and Zhang, 2014)	√	√	Pixel-level
FSDAF (Zhu et al., 2016)	√	√	Pixel-level
HSTAFM (Chen et al., 2016)	√	√	Pixel- and feature-level
Fit-FC (Wang and Atkinson, 2017)	√	√	Pixel-level